Evaluating the Performance of Nonnegative Matrix Factorization for Constructing Semantic Spaces: Comparison to Latent Semantic Analysis

Akira Utsumi
Department of Informatics
The University of Electro-Communications
Chofushi, Tokyo 182-8585, Japan
utsumi@se.uec.ac.jp

***Abstract*—This study examines the ability of nonnegative matrix factorization (NMF) as a method for constructing semantic spaces, in which the meaning of each word is represented by a high-dimensional vector. The performance of two tests (i.e., a multiple-choice synonym test and a word association test) is compared between NMF and latent semantic analysis (LSA), which is the most popular method for constructing semantic spaces. As a result, it was found that NMF did not outperform LSA in either test. This finding indicates that NMF is less effective in acquiring word meanings than expected in the literature; in other words, the finding provides evidence for the ability of LSA to represent semantic meanings. Some properties of NMF were also revealed with reference to its ability to represent word meanings; the random initialization was superior to the SVD-based initialization, and the Euclidean distance is more appropriate for the objective function of NMF than the KL-divergence. In addition, it was shown that the inner product was a more appropriate method for measuring the syntagmatic similarity in a semantic space model, while the cosine was a better method for computing the paradigmatic similarity.**

***Index Terms*—Semantic space, Word meaning, Nonnegative matrix factorization, Latent semantic analysis, Singular value decomposition, Word association, Syntagmatic and paradigmatic relations**

I. Introduction

Recent research effort in computational lexical semantics has been directed at a semantic space model [1]–[3], a corpus-based method for acquiring and representing the meaning of words. Semantic space models are computationally efficient as a way of representing meanings of words, because they take much less time and less effort to construct meaning representation and they can provide a more fine-grained similarity measure than other representation methods such as thesauri (e.g., WordNet). Semantic space models are also psychologically plausible; a number of studies have shown that a semantic space achieves remarkably good performance for simulating human verbal behavior [1], [4].

In a semantic space, the meaning of each word is represented by a high-dimensional vector, that is, *a word vector*. The degree of semantic similarity between any two words can be easily computed as, for example, the cosine of the angle formed by their word vectors. For example, using a semantic space of Japanese words used in this paper, the cosine similarity between *PC* (“*pasokon*” in Japanese; personal computer) and *Windows* (“*uindouzu*” in Japanese; Microsoft’s OS) is computed as 0.77, while the cosine similarity between *PC* and *window* (“*mado*” in Japanese; glass in the wall) is computed as 0.13.

Semantic spaces (or word vectors) are constructed from large bodies of text by observing distributional statistics of word occurrence. The method for constructing semantic spaces generally comprises the following two steps:

1) *Matrix construction*: All n content words are represented as m dimensional feature vectors, and an $n \times m$ matrix M is constructed using n word vectors as rows.
2) *Dimensionality reduction*: The dimension of M’s row vectors is reduced from the initial dimension m to d.

As a result, a d-dimensional semantic space including n words is generated. Latent semantic analysis [1], [5] (henceforth, LSA) is the most popular method for constructing semantic spaces; it uses the frequency of words in a document to compute initial vectors, and singular value decomposition (henceforth, SVD) to reduce the dimension. LSA has been applied to many language-related tasks — e.g., information retrieval [6], [7], essay scoring [8], noun compound comprehension [9], and design communication [10] — and its performance has been well evaluated.

Among dimensionality reduction techniques, nonnegative matrix factorization (henceforth, NMF) [11], [12] has been widely used in recent years. NMF has proved highly useful for diverse fields of science such as bioinformatics [13], face and image recognition [14], [15], color science [16], document clustering [17], and automatic summarization [18], among others. However, no studies have ever tried to examine the applicability of NMF to semantic space construction. This is particularly surprising given that Lee and Seung, who first proposed NMF, suggested that NMF can generate semantic spaces in their seminal paper [11]. It is interesting and vital for research on both semantic space models and NMF to evaluate the representational ability of semantic spaces generated by NMF.

Therefore, the purpose of this paper is to evaluate the ability of NMF for constructing appropriate semantic spaces.

978-1-4244-6587-3/10/$25.00 É2010 IEEE

In particular, NMF-based semantic spaces are compared to LSA spaces in terms of their performance in two tests, namely a multiple-choice synonym test and a word association test. These tests have been used as a performance measure by many studies on the semantic space model.

II. Constructing Semantic Spaces

As mentioned in the introduction, semantic spaces are constructed in the two steps, namely matrix construction and dimensionality reduction.

A. Matrix construction

In LSA, all content words w_i in a corpus are represented as m-dimensional feature vectors $\boldsymbol{w}_i$ that are computed as follows:

$$\begin{aligned} \boldsymbol{w}_i &= (w_{i1}, w_{i2}, \cdots, w_{im}) && (1) \\ w_{ij} &= tf_{ij} \times \left(1 + \frac{\sum_{k=1}^{m} P_{ik} \log P_{ik}}{\log m}\right) && (2) \\ P_{ij} &= \frac{tf_{ij}}{\sum_{k=1}^{m} tf_{ik}}, && (3) \end{aligned}$$

where tf_{ij} is the frequency of the i-th word w_i in the j-th document (i.e., the number of times the word w_i occurs in the j-th document). A word-document matrix M is then constructed using n feature vectors $\boldsymbol{w}_i$ as rows. This word-document matrix is common to all semantic spaces that are compared in the evaluation experiment of this study.

B. Dimensionality reduction

1) SVD: LSA uses SVD as a dimensionality reduction technique. SVD factorizes a word-document matrix M as the product of three matrices

$$M = U \Sigma V^T, \tag{4}$$

where the diagonal matrix Σ consists of r singular values that are arranged in nonincreasing order such that r is the rank of M. When we use a $d \times d$ matrix Σ_d consisting of the largest d singular values, the matrix M is approximated by $U_d \Sigma_d V_d^T$, where U_d is an $n \times d$ matrix consisting of the corresponding d left singular vectors as columns and V_d is an $m \times d$ matrix consisting of the corresponding d right singular vectors as columns. The i-th row of $U_d \Sigma_d$ corresponds to a d-dimensional "reduced word vector" for the word w_i.

2) NMF: NMF approximates the matrix M by the product of an $n \times d$ nonnegative matrix W and a $d \times m$ nonnegative matrix H.

$$M \approx WH \tag{5}$$

The i-th row of the matrix W corresponds to a d-dimensional vector for the word w_i.

The following cost functions are widely used to find an approximate factorization (5).

$$\begin{aligned} U(M, WH) &= ||M - WH||^2 = \sum_{ij} (M_{ij} - (WH)_{ij})^2 && (6) \\ D(M, WH) &= \sum_{ij} \left(-M_{ij} \log (WH)_{ij} + (WH)_{ij}\right) && (7) \end{aligned}$$

Equation (6) is simply the square of the Euclidean distance between M and WH, while Equation (7) denotes a generalized Kullback-Leibler divergence (henceforth, KL-divergence). In this paper, both functions are used for evaluation.

For these objective functions to converge to a local minimum, the following multiplicative update rules are employed [12].

- Update rules for Euclidean distance (6)

$$\begin{aligned} W'_{ij} &\leftarrow W_{ij} \frac{(MH^T)_{ij}}{(WHH^T)_{ij}} && (8) \\ W_{ij} &\leftarrow \frac{W'_{ij}}{\sum_k W'_{ik}} \\ H_{ij} &\leftarrow H_{ij} \frac{(W^T M)_{ij}}{(W^T W H)_{ij}} && (9) \end{aligned}$$

- Update rules for KL-divergence (7)

$$\begin{aligned} W'_{ij} &\leftarrow W_{ij} \sum_k \frac{M_{ik}}{(WH)_{ik}} H_{jk} && (10) \\ W_{ij} &\leftarrow \frac{W'_{ij}}{\sum_k W'_{ik}} \\ H_{ij} &\leftarrow H_{ij} \sum_k W_{ki} \frac{M_{kj}}{(WH)_{kj}} && (11) \end{aligned}$$

By computing iteratively two matrices W and H using the update rules, we can obtain a locally optimal approximation.

One important issue that must be addressed for the iterative computation of NMF is how two matrices W and H are initialized. Most NMF studies use random nonnegative initialization, but some studies have proposed an initialization algorithm that contains no randomization. This study compares two initialization methods, a random initialization (W and H are initialized as random values between 0 and 1) and a SVD-based initialization [19]. In this SVD-based initialization algorithm, W and H are initialized as $U_d\sqrt{\Sigma_d}$ and $\sqrt{\Sigma_d}V_d^T$. To guarantee the nonnegativity of the matrices, only positive elements or only negative elements are used for initialization. Which element is selected is determined depending on their norm. (For the details of the SVD-based initialization method, see [19]).

C. Corpus

As a corpus from which semantic spaces are constructed, this study uses 8,451 Japanese Mainichi newspaper articles (published in 1998 and 1999) that contain more than or equal to 229 content words. This corpus contains 34,901 different words that occur 10 times or more in the articles. Hence, the dimensions of the matrix M are $n = 34901$ and $m = 8451$.

III. Experiment 1: Multiple-Choice Synonym Test

A. Method

In order to compare different semantic spaces in terms of the ability to judge the semantic similarity between words, this study conducted a multiple-choice synonym test. Each item of a synonym test comprises a stem word and five alternative

TABLE I
EXAMPLES OF THE MULTIPLE-CHOICE SYNONYM TEST

Stem word	Alternative words
author (*chosha*)	writer (*sakusha*), prince (*koutaishi*), war (*senso*), impulse (*shoudou*), both sides (*ryomen*)
important (*juyo*)	precious (*kicho*), effortless (*raku*), impossible (*fukanou*), happy (*happii*), unconscious (*muishiki*)

Note. Underlined words denote correct answers. The original Japanese words used in the test are shown in parentheses preceded by their English translations.

TABLE II
THE MAXIMUM ACCURACY OF THE LSA AND FOUR NMF SPACES

LSA	NMF			
	Euclidean distance		KL-divergence	
	random	svd	random	svd
57.76 (600)	52.96 (300)	49.80 (300)	50.26 (300)	50.10 (200)

Note. Numbers in parentheses denote the number of dimension d that achieved the maximum accuracy.

words from which the test-taker is asked to choose one with the most similar meaning to the stem word.

In this study, test items were automatically generated using a Japanese thesaurus "Nihongo Dai-Thesaurus" [20]. This thesaurus consists of 1,044 basic categories, which are divided into nearly 14,000 semantic categories. The thesaurus contains nearly 200,000 words, most of which are classified into multiple semantic categories. Test items were generated as follows. First, all words that occur 100 times or more in the corpus were chosen as stem words. Second, for each of the stem words, one correct alternative word was chosen randomly from the semantic categories that included the stem word. Finally, other four alternatives were chosen randomly from the basic categories that did not include either the stem word or the correct answer. All alternative words were chosen such that they occurred 50 times or more in the corpus. As a result, 3,918 test items were generated, some examples of which are shown in Table I.

In the experiment, five semantic spaces — i.e., one LSA (SVD) space and four NMF spaces constructed by two objective functions and two initialization methods — were compared in terms of the percentage of correct answers (i.e., accuracy). The computer's choice using semantic spaces was determined by computing the cosine similarity between the stem word and each of the five alternative words and choosing the word with the highest similarity.

B. Result and discussion

For each of the five semantic spaces, the multiple-choice synonym test described above was conducted and the percentage of correct choices was calculated. The dimension d of the semantic spaces varied between 100 and 1000 in steps of 100.

1) Comparison between LSA and NMF: Fig. 1 shows the accuracy (i.e., the percentage of correct answers) of the LSA and NMF spaces. The results of NMF spaces shown in

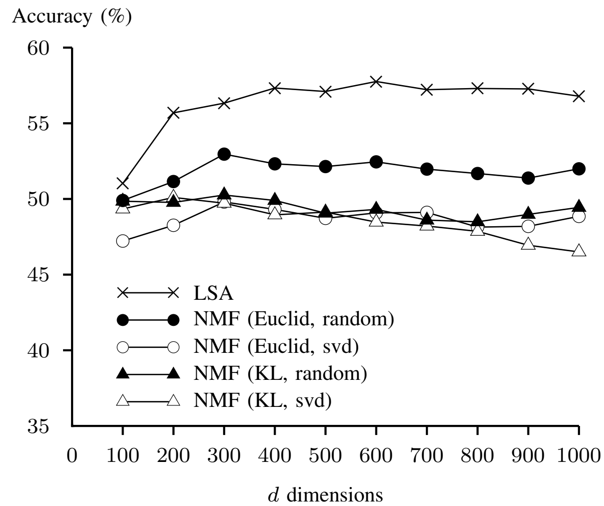


Fig. 1. Accuracy of synonym tests for LSA and NMF semantic spaces

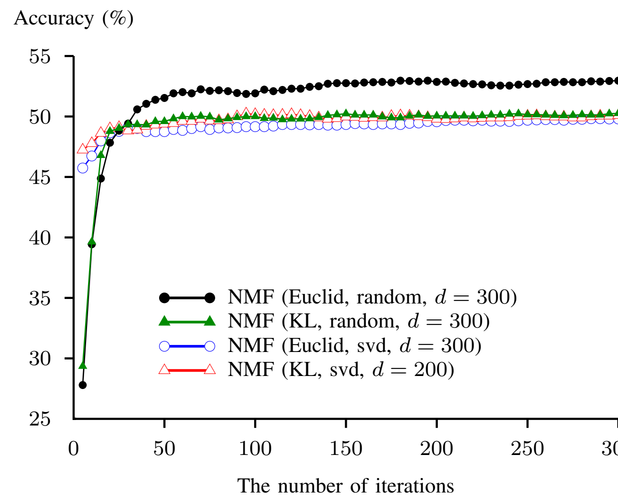


Fig. 2. Accuracy of synonym tests of NMF spaces as a function of the number of iterations

the figure are after 300 iterations, before which the NMF computation converged.[1] Table II lists the maximum accuracy for the LSA and NMF spaces. Note that "Euclid" and "KL" in Fig. 1 denote that Euclidean distance and KL-divergence are used as an objective function, and "random" and "svd" denote the random and svd-based initialization.

The most important result shown in Fig. 1 and Table II is that, regardless of objective function and initialization method, the NMF-based semantic spaces did not outperform the LSA-based semantic space. This result was replicated when the accuracy of NMF spaces was computed before 300 iterations, as shown in Fig. 2. Fig. 2 demonstrates that the accuracy of NMF spaces converged until 100 iterations, regardless of objective function and initialization method. These findings

[1]The NMF-based semantic spaces were computed using CUDA on an Intel Xeon CPU and Tesla C1060 GPU. The computation time for 300 iterations was about 50 minutes when the number of dimensions $d = 100$, and thus eight hours and twenty minutes when $d = 1000$.

suggest that NMF may be less appropriate for generating semantic spaces than SVD, which is used by LSA.

2) Effect of the objective function and the initialization method: Another important result of Fig. 1 (and Table II) is that the NMF spaces computed by optimizing the Euclidean distance (6) outperformed the NMF spaces obtained by optimizing the KL-divergence (7), especially when matrices are initialized randomly. This may suggest that the Euclidean objective function is more appropriate for constructing semantic spaces than the KL-divergence objective. In addition, Fig. 1 and Table II reveal that the random initialization yielded better performance than the SVD-based initialization; the SVD-based initialization was not effective in improving the quality of semantic spaces, although it led to faster convergence as shown in Fig. 2.

One concern with the random initialization is that the performance of NMF-based semantic spaces may depend on the initial W and H values and thus only one solution cannot seem to obtain accurate (or best) performance. [2] Hence, to examine the extent to which the accuracy of the multiple-choice synonym test varied depending on random initializations, I generated 20 different semantic spaces ($d = 100$) with the Euclidean objective function using different random initializations, and computed the accuracy of these spaces after 100 iterations. The mean accuracy of these 20 spaces was 49.82% (SD=0.49), and the maximum and minimum accuracies were 50.59% and 48.65%. The 95 percent confidence interval was 49.59% to 50.05% and very narrow, thus indicating that the random initialization does not have a serious influence on the overall performance of NMF-based semantic spaces.

3) Effect of similarity measure: The inferiority of NMF-based semantic spaces found in this experiment may be due to the cosine similarity used as a measure of the semantic relatedness between words. The semantic space obtained by the NMF algorithm can be regarded as the topic model [21] in that the columns of W represent semantic features or topics [11]. The relation between NMF and the topic model can also be justified by the finding [22] on the equivalence between NMF and PLSI [23], which is a probabilistic version of LSA and can be regarded as a version of the topic model [21]. In the topic model, each word is represented as a "local representation," i.e., a mixture of probabilistic topics, and thus the cosine may not be appropriate for computing the semantic relatedness in the topic model.

Hence, the multiple-choice synonym test was conducted by computing the similarity between words using two alternative measures, namely the inner product and the (approximated) conditional probability of one word given another, which are shown to be effective in a few applications of the topic model or the semantic space model [21]. (The details of computing the conditional probability for NMF spaces are given in the appendix.) Fig. 3 shows the result of the synonym test for the inner product and the conditional probability, together with the result for the cosine. (Note that the conditional probability was not applied to LSA spaces because their word vectors have negative values. All NMF spaces in Fig. 3 were generated by the random initialization method.) Again, NMF did not outperform LSA; both the inner product and the probability gave worse performance than the cosine. In particular, the inner product considerably degraded the accuracy, possibly because the inner product is affected by word frequency (i.e., frequent words have a higher inner product), but word frequency is not an important component in choosing a correct answer.

[2]The best way to cope with this problem is to run several instances of the NMF algorithm using different random initializations and then select the best solution (or compute the average performance). However, NMF computation generally takes much time and thus the overall experimental process becomes quite expensive. It becomes more serious when a large matrix such as a term-document matrix in this study should be factorized.

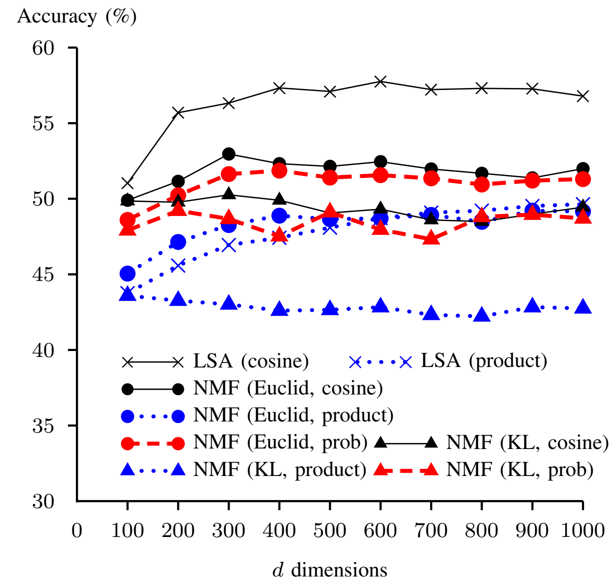


Fig. 3. Accuracy of synonym tests computed by different similarity measures

IV. Experiment 2: Word Association Test

A. Method

In order to compare the ability of the semantic spaces to predict word association, this study conducted a word association test, in which 1150 human word association pairs were used. They were chosen from the top 10 associates of 201 stimulus words in a Japanese word association norm "Renso Kijunhyo" [24]. For example, the stimulus word *writer* (*sakka*) is associated with *novel* (*shosetsu*), *book* (*hon*), *author* (*sakusha*), *work* (*sakuhin*), and *write* (*kaku*).

In the word association test, for each pair (t_i^S, t_i^A) of a stimulus word t_i^S and an associate word t_i^A, the rank r_i of t_i^A is assessed by computing the similarity between t_i^S and all other words in a semantic space (in this study, $n - 1 = 34900$ words) and sorting all words in descending order of the similarity. The performance of the semantic spaces in predicting word association is measured by the mean recall $R_{ave}(T)$ of

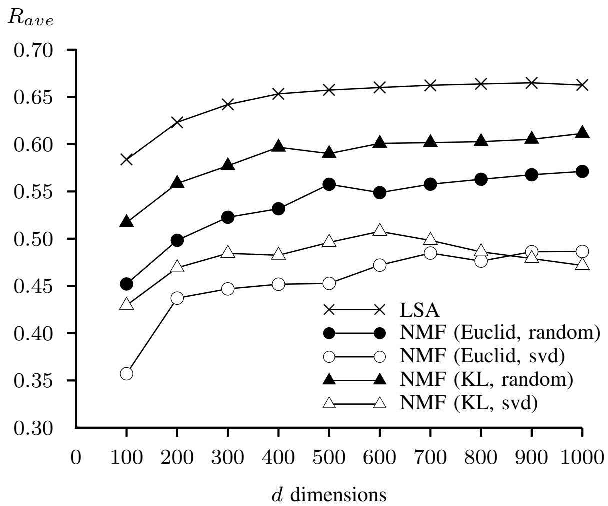


Fig. 4. Mean recall of the word association test for LSA and NMF spaces

associates in a set of word association pairs T.

$$R_{ave}(T) = \frac{\sum_{i=1}^{\lfloor n/1000 \rfloor} R_{100\times i}(T)}{\lfloor n/1000 \rfloor} = \frac{\sum_{i=1}^{34} R_{100\times i}(T)}{34} \quad (12)$$

$$R_i(T) = \frac{\mid \{(t_j^S, t_j^A) \mid r_j \leq i\} \mid}{|T|} \quad (13)$$

The recall $R_i(T)$ is calculated as the fraction of associates which are included in the set of the top i words with the highest similarity to the stimulus words. Hence, the mean recall R_{ave} is the average of R_{100}, R_{200}, $\cdots$, and R_{3400}, which roughly evaluates the percentage of human associate words that are ranked high on the word list sorted by a semantic space. (R_{3400} denotes the percentage of human associates that are included in the top ten percent on the list.)

B. Result and discussion

Fig. 4 shows the mean recall of the associates predicted by the LSA space and the four NMF spaces. For the NMF spaces, mean recall was computed after 300 iterations. Fig. 5 shows the mean recall of the LSA and NMF spaces (with the random initialization) for different similarity measures, i.e., the cosine, the inner product, and the probability.

1) Comparison between NMF and LSA: As can be seen from Fig. 4, the NMF spaces did not outperform the LSA space regardless of objective function and initialization method, which is consistent with the result of the multiple-choice synonym test. These convergent findings indicate that NMF is less effective in acquiring word meanings than expected in the literature; in other words, LSA proves highly appropriate for representing the meaning of words.

Fig. 5 shows that NMF did not outperform LSA even when either the inner product or the probability was used as a similarity measure. The probability drastically degraded the performance of NMF spaces. The inner product improved the performance of NMF spaces, which was comparable to the

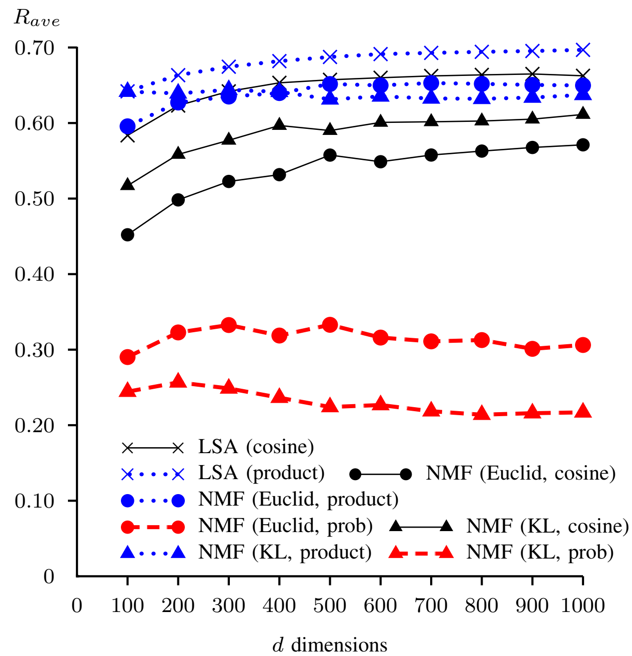


Fig. 5. Mean recall of the word association test for different similarity measures

performance of the LSA space with the cosine but was worse than that of the LSA space with the inner product.

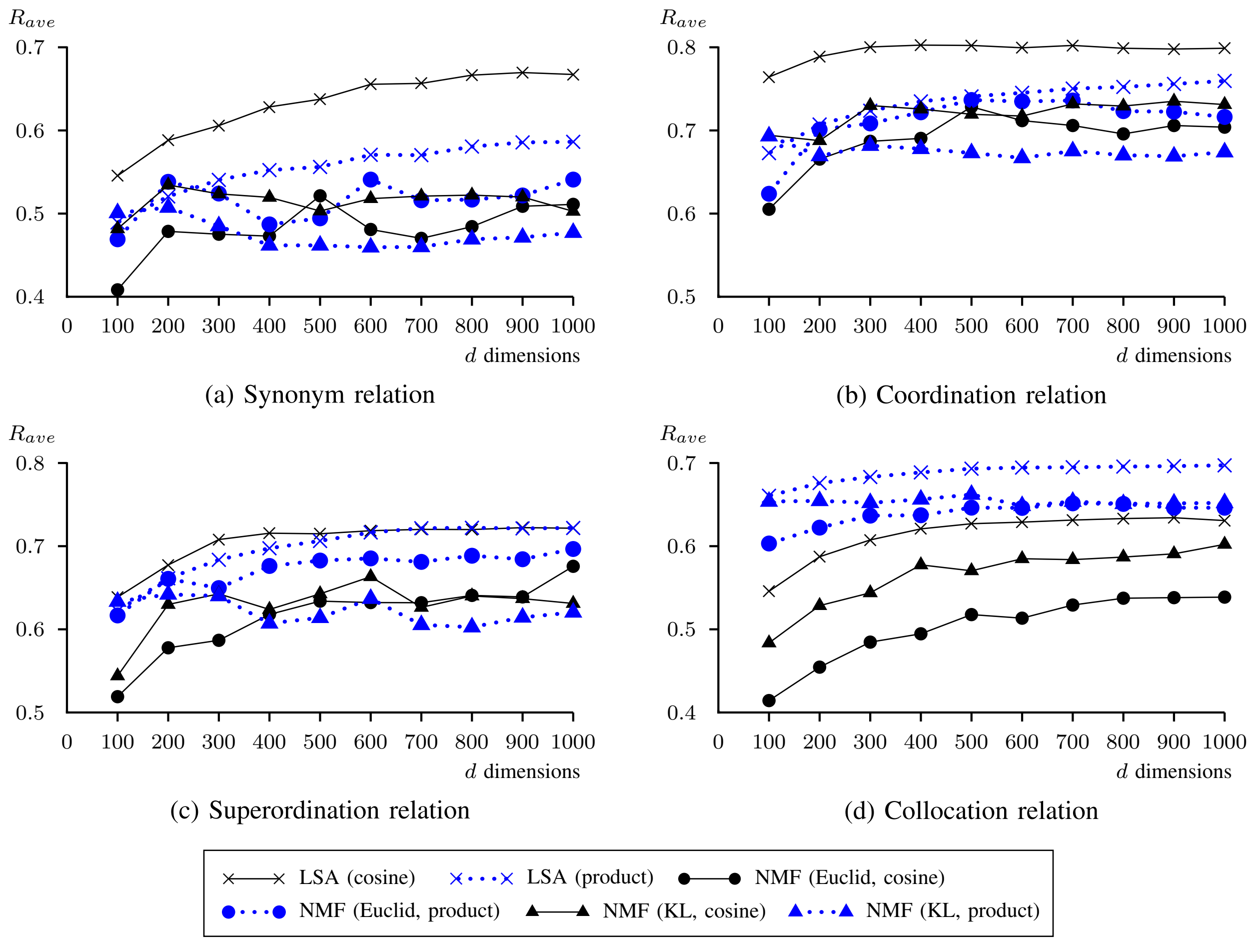
2) Effect of the objective function and the initialization method: As shown in Fig. 4, the SVD-based initialization gave much worse performance than the random initialization, which is consistent with the result of the multiple-choice synonym test. On the other hand, the word association test yielded one different result; in contrast to the result of the synonym test, the NMF spaces with the KL-divergence achieved better performance in predicting word association than the NMF spaces with the Euclidean distance. This result seems to suggest that the objective function appropriate for constructing a semantic space depends on the task to be performed by the semantic space. However, Fig. 5 shows that, when the inner product was used as a similarity measure, the NMF space with the Euclidean distance gave better performance. Hence, it may be concluded that the Euclidean distance generally generates better semantic spaces than the KL-divergence.

3) Effect of similarity measure: As shown in Fig. 5, although the probability drastically degraded the performance of NMF spaces, the inner product improved the performance of both the LSA and NMF spaces as compared to the cosine. This result differs from the result of the multiple-choice synonym test that the inner product degraded the performance; such the positive effect of the inner product would be due to the property of word association that word frequency is an important determinant of whether a word will be named as an associate.

4) Relation between semantic spaces and semantic relations: Word association is based on several semantic relations. For example, *author* is synonymously related to *writer*, while *novel* is functionally related to *writer* in that a writer is

TABLE III
FOUR TYPES OF SEMANTIC RELATIONS THAT UNDERLIE WORD ASSOCIATION

Relation	Definition and description	Examples
Synonym	Two words have identical or very similar meanings.	student – pupil, buy – purchase
Coordination	Two words (including antonyms) cluster together on the same level of detail.	desk – chair, black – white (antonymy)
Superordination	One word is a superordinate (i.e., hypernym) of another word. This category includes holonymy.	animal – dog, color – red, car – engine (holonymy)
Collocation	Two words are likely to cooccur in the text, because they form a predicate-argument structure.	rose – red, love – affair, baseball – play



(a) Synonym relation

(b) Coordination relation

(c) Superordination relation

(d) Collocation relation

Fig. 6. Mean recall for each of the four semantic relations

usually engaged in writing novels. Semantic relations are generally classified into four types shown in Table III [25]. Some studies have shown that different semantic spaces well represent different semantic relations [26], [27].

Therefore, in this study, the word association pairs (i.e., stimulus and associate words) were classified by their underlying semantic relations, [3] and semantic spaces were evaluated for each of these semantic relations. Fig. 6 shows the mean recall of the LSA and NMF spaces calculated for each semantic relation. (The results of NMF spaces for the probability are not shown because their mean recalls are much lower, as shown in Fig. 5.) Table IV shows the maximum mean recall for the LSA and NMF spaces.

Fig. 6 demonstrates that again NMF spaces did not reach the level of the LSA performance for either relation, but their degrees were different. The inner product did not improve the performance for synonym and coordination relations (which are often referred to as *paradigmatic relations*), but improved the performance for superordination and collocation relations (which are often referred to as *syntagmatic relations*). This difference may be caused by the different effect of word

[3] As a result of classification, 77, 163, 127 and 783 association pairs were classified into synonym, coordination, superordination and collocation, respectively. Note that about 68% of word association pairs are based on the collocation relation.

TABLE IV
THE MAXIMUM MEAN RECALL OF THE LSA AND FOUR NMF SPACES

	LSA		NMF			
			Euclid		KL	
	cos	dot	cos	dot	cos	dot
Synonym	.670 (900)	.586 (1000)	.521 (500)	.541 (600)	.534 (200)	.507 (200)
Coordination	.803 (400)	.760 (1000)	.728 (500)	.737 (500)	.735 (900)	.693 (100)
Superordination	.722 (900)	.722 (800)	.676 (1000)	.697 (1000)	.663 (600)	.642 (200)
Collocation	.634 (900)	.697 (1000)	.539 (1000)	.652 (700)	.602 (1000)	.662 (500)
All	.665 (900)	.697 (1000)	.571 (1000)	.653 (700)	.611 (1000)	.644 (300)

Note. Numbers in parentheses denote the number of dimension d that achieved the maximum mean recall. cos = cosine. dot = inner product.

frequency on human judgment of semantic relatedness. Judgment on whether two words are paradigmatically related is not significantly affected by word frequency. For example, *buy* and *purchase* are easily judged as synonymous regardless of their frequencies. On the other hand, people seem to judge whether two words are syntagmatically related depending on their frequency; for example, *apple* is easily associated with *fruit*, but *carambola* (i.e., star-fruit) is much less likely to be associated with *fruit*. Hence, the effect of word frequency on word association shown in Fig. 5 can be attributed to the property of syntagmatic relations, because many word association pairs (about 79%) are based on syntagmatic relations.

Whether a better NMF space can be generated by the Euclidean distance or by the KL-divergence also depended on semantic relations. Word association pairs based on the synonym, coordination, or superordination relations were better predicted by the NMF space with the Euclidean distance and the inner product. Table IV shows that for these relations the NMF space with the Euclidean distance and the inner product yielded the maximum mean recall among NMF spaces, although not clearly shown in Figure 6. On the other hand, word association pairs based on the collocation relation were better predicted by the NMF spaces with the KL-divergence regardless of their similarity measure, as shown in Figure 6 and Table IV.

V. CONCLUSION

Through two evaluation experiments for LSA and NMF semantic spaces, the following findings are obtained:

- LSA yielded better performance in both synonym judgment and word association than NMF. This finding indicates that NMF is less effective in generating semantic spaces; in other words, SVD is superior to NMF as a dimensionality reduction technique for the semantic space model.
- A little surprisingly, the random initialization yielded better performance in generating NMF-based semantic spaces than the SVD-based initialization.
- Overall, the Euclidean distance generates better semantic spaces than the KL-divergence, but the KL-divergence is more appropriate for representing the collocation relation.
- The inner product is a more appropriate method for measuring the syntagmatic similarity (i.e., the semantic relatedness that depends on word frequency) in a semantic space model, while the cosine is more appropriate for computing the paradigmatic similarity (i.e., the semantic similarity that is not affected by word frequency). The inner product improves the quality of NMF spaces to a greater extent than that of LSA spaces.

I am now extending this work to examine in more detail a potential ability of NMF to generate high-quality semantic spaces. In addition, I am trying to develop a novel initialization method for NMF that generates better semantic spaces. It would also be vital for further research to develop a more powerful method for generating semantic spaces.

ACKNOWLEDGMENT

This study was supported by a Grant-in-Aid for Scientific Research C (No.20500234) from Japan Society for the Promotion of Science.

APPENDIX

The conditional probability $P(w_i|w_j)$ of a word w_i given another word w_j can be approximately computed under the assumption that the k-th element W_{ik} of the word vector w_i represents the probability $P(w_i|t_k)$ of w_i given the k-th topic t_k and a prior probability over topics is uniform, i.e., $P(t_k) = 1/d$.

$$
\begin{aligned}
P(w_i|w_j) &= \sum_{k=1}^{d} P(w_i|t_k)P(t_k|w_j) \\
&= \sum_{k=1}^{d} P(w_i|t_k) \cdot \frac{P(w_j|t_k)P(t_k)}{P(w_j)} \\
&= \sum_{k=1}^{d} P(w_i|t_k) \cdot \frac{P(w_j|t_k)P(t_k)}{\sum_l P(w_j|t_l)P(t_l)} \\
&= \frac{\sum_{k=1}^{d} W_{ik}W_{jk}}{\sum_{k=1}^{d} W_{jk}} \qquad (14)
\end{aligned}
$$

Therefore, the approximated probability is equal to the inner product divided by the L_1 norm of the vector of w_j, because elements of word vectors W_{jk} are nonnegative.

REFERENCES

[1] T. Landauer, D. McNamara, S. Dennis, and W. Kintsch, *Handbook of Latent Semantic Analysis*. Lawrence Erlbaum Associates, 2007.
[2] S. Padó and M. Lapata, "Dependency-based construction of semantic space models," *Computational Linguistics*, vol. 33, no. 2, pp. 161–199, 2007.
[3] H. Schütze, "Automatic word sense discrimination," *Computational Linguistics*, vol. 24, no. 1, pp. 97–123, 1998.
[4] J. Bullinaria and J. Levy, "Extracting semantic representations from word co-occurrence statistics: A computational study," *Behavior Research Methods*, vol. 39, no. 3, pp. 510–526, 2007.

[5] T. K. Landauer and S. T. Dumais, "A solution to Plato's problem: The latent semantic analysis theory of the acquisition, induction, and representation of knowledge," *Psychological Review*, vol. 104, pp. 211–240, 1997.
[6] S. Deerwester, S. T. Dumais, G. W. Furnas, T. L. Landauer, and R. Harshman, "Indexing by latent semantic analysis," *Journal of the American Society For Information Science*, vol. 41, no. 6, pp. 391–407, 1990.
[7] D. Widdows, *Geometry and Meaning*. CSLI Publications, 2004.
[8] T. K. Landauer, D. Laham, and P. W. Foltz, "Automated scoring and annotation of essays with the Intelligent Essay Assessor," in *Automated Essay scoring: A Cross-disciplinary Perspective*, M. Shermis and J. Burstein, Eds. Mahwah, NJ: Lawrence Erlbaum Associates, 2003, pp. 87–112.
[9] A. Utsumi, "Computational semantics of noun compounds in a semantic space model," in *Proceedings of the 21st International Joint Conference on Artificial Intelligence (IJCAI-09)*, 2009, pp. 1568–1573.
[10] A. Dong, "The latent semantic approach to studying design team communication," *Design Studies*, vol. 26, pp. 445–461, 2005.
[11] D. D. Lee and S. Seung, "Learning the parts of objects by non-negative matrix factorization," *Nature*, pp. 788–791, 1999.
[12] ——, "Algorithms for non-negative matrix factorization," *Advances in Neural Information Processing Systems*, vol. 13, pp. 556–562, 2000.
[13] J. Brunet, P. Tamayo, T. Golub, and J. Mesirov, "Metagenes and molecular pattern discovery using matrix factorization," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 101, no. 12, pp. 4164–4169, 2004.
[14] D. Guillamet, J. Vitria, and B. Schiele, "Introducing a weighted non-negative matrix factorization for image classification," *Pattern Recognition Letters*, vol. 24, pp. 2447–2454, 2003.
[15] S. Zafeiriou, A. Tefas, I. Buciu, and I. Pitas, "Exploiting discriminant information in nonnegative matrix factorization with application to frontal face verification," *IEEE Trans. Neural Networks*, vol. 17, no. 3, pp. 683–695, 2006.
[16] G. Buchsbaum and O. Bloch, "Color categories revealed by non-negative matrix factorization of Munsell color spectra," *Vision Research*, vol. 42, no. 5, pp. 559–563, 2002.
[17] F. Shahnaz, M. Berry, V. Pauca, and R. Plemmons, "Document clustering using nonnegative matrix factorization," *Information Processing and Management*, vol. 42, pp. 373–386, 2006.
[18] J. Lee, S. Park, C. Ahn, and D. Kim, "Automatic generic document summarization based on non-negative matrix factorization," *Information Processing and Management*, vol. 45, pp. 20–34, 2009.
[19] C. Boutsidis and E. Gallopoulos, "SVD based initialization: A head start for nonnegative matrix factorization," *Pattern Recognition*, vol. 41, pp. 1350–1362, 2008.
[20] T. Yamaguchi, *Nihongo Dai-Thesaurus CD-ROM*. Taishukan Shoten, Tokyo, 2006.
[21] T. Griffiths, M. Steyvers, and J. Tenenbaum, "Topics in semantic representation," *Psychological Review*, vol. 114, pp. 211–244, 2007.
[22] C. Ding, T. Li, and W. Peng, "On the equivalence between non-negative matrix factorization and probabilistic latent semantic indexing," *Computational Statistics and Data Analysis*, vol. 52, pp. 3913–3927, 2008.
[23] T. Hofmann, "Probabilistic latent semantic indexing," in *Proceedings of the 22nd International Conference on Research and Development in Information Retrieval (SIGIR'99)*, 1999, pp. 50–57.
[24] T. Umemoto, *Renso Kijunhyo (Free Association Norm)*. Tokyo Daigaku Shuppankai, Tokyo, 1969.
[25] J. Aitchison, *Words in the Mind: An Introduction to the Mental Lexicon, 3rd Edition*. Oxford, Basil Blackwell, 2003.
[26] M. Sahlgren, "The word-space model: Using distributional analysis to represent syntagmatic and paradigmatic relations between words in high-dimensional vector spaces," Ph.D. dissertation, Stockholm University, 2006.
[27] A. Utsumi, "Exploring the relationship between semantic spaces and semantic relations," in *Proceedings of the 7th International Joint Conference on Language Resources and Evaluation (LREC2010)*, 2010, pp. 257–262.